变分法简介

蔡家麒

华中科技大学

caidish@hust.edu.cn

November 2, 2016

Overview

1. 线性代数
   - 另一个角度看待二次型（厄米型）的主轴变换
   - 本征值的极值性
2. 初始变分法
   - 初始变分法证明本征值的极值性
   - 本征值的极小极大值性
3. 形式变分法
   - 初始变分方法解决泛函极值的必要条件
   - 通过形式变分法计算简单的泛函极值问题
   - 有附加约束的欧拉方程的导出
4. 直接变分方法
   - Ritz 方法
   - 有限差分法
   - 直接方法的应用举例

主轴变换

Theorem (定义)

把二次型 $K(x) = \sum_{q,p=1}^{n} x_p k_{pq} x_q$

化作平方和 $K(x) = \sum_{p=1}^{n} \kappa_p y_p^2$

的正交变换 $x_p = \sum_{q=1}^{n} y_q l_{qp}$

这是一个只做旋转和反演的线性变换，我们把这定义为叫做主轴变换。

这类主轴变换在代数中极为重要，我们先将上述表述写成矩阵的形式，并用内积表出。

主轴变换

设

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

那么 $K(x) = x^T K x$ 变换到 $K(x) = y^T \kappa y$
其中 κ 是一个对角矩阵
相应的正交变换是

$$x = Ly$$

其中

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

利用下面的定理

Theorem (魏尔斯特拉斯定理)

对于有限闭区域上的多元连续函数，总在该区域内取得极大值。

我们首先取一个归一化的约束 $x^T x = 1$，他的几何意义非常显然，是自然坐标下的一个单位球面。
利用上述定理，我们总能找到一个矢量

$$l_1 = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1n} \end{pmatrix}$$

使得在 $x^T = l_1$ 时，$K(x) = x^T K x$ 取得极大值 κ_1

此外，由于空间是 n 维的，我们在空间中也总能找到一个和 l_1 正交的向量 l_2,$x^T = l_2$, 使得 $K(x) = x^T Kx$ 满足球面约束的情况下取得极大值 κ_2。

正交条件和约束条件的几何意义是十分的明显的，就是单位球和垂直于 l_1 的" 超平面" 相交而成的一个" 超曲线"。

此外，我们可以一直这样扩充, 让 l_m 总是垂直于 $l_1, l_2, ..., l_m - 1$,$x^T = l_m$, 使得 $K(x) = x^T Kx$ 满足球面约束的情况下取得极大值 κ_m。
这样一直扩充到了 l_n，这个 n 个向量称作本征矢量，马上就会说明原因。这样的 l_n 排成一列构成的矩阵，约束条件 $x^T x = 1$ 和正交约束条件保证了这是一个正交矩阵，由这个矩阵生成的正交变换

$$x = Ly$$

就是我们要求的答案，这个结论在二维、三维维情况下是十分显然的（以椭球为例子），但是对于高维，我们需要一个简单的证明（不赘述）。

本征值的极值性

我们回顾一下，通常对于二次型的主轴变换的求法是:

- 我们要求的是一个正交矩阵，将 K 通过正交变换（等价于相似变换）使他成为对角型;
- 事实上，我们利用了相似变换的性质，求得了本征系统，设其中的本征值是 λ_n, 对应的本征矢量是 ψ_n
- 这些本征值就是变换之后的对角矩阵的对角元, 对应的本征矢量排成列（排成列和后面对应），可以得到 $y \to x$ 的线性变换。

对比这里求主轴变换的方法和上面对主轴变换的定义，事实上

$$\lambda_n := \kappa_n$$

$$\psi_n := l_n$$

不需要证明，本征值的极值性已经体现的非常明显。这一套理论可以推广到复数域，只需要把厄米型定义成

$$K(x, \bar{x}) = \bar{x} K x$$

本征值的极值性

Theorem (本征值的极值性)

设 V 是 C 上的一个酉空间，H 是其上的一个厄米变换，设 H 的谱有下限的，并且各个本征值及其排列如下：$\lambda_0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \dots.$，则归一化的厄米型 $\frac{(\alpha,H\alpha)}{(\alpha,\alpha)} = \left(\frac{\alpha}{|\alpha|}, H\frac{\alpha}{|\alpha|}\right)$ 的极小值是：

- λ_0，*倘若 α 是 V 中任意的一个非零向量*
- λ_1，*倘若 α 是 V 中任意的一个非零向量，满足约束 $(\alpha_0, \alpha) = 0$*
- λ_m，*倘若 α 是 V 中任意的一个非零向量，满足约束* $(\alpha_0, \alpha) = (\alpha_1, \alpha) = \dots. = (\alpha_{m-1}, \alpha)$

事实上，这里的厄米型是我们解决的第一个泛函的极值问题，
$J[\alpha] = \frac{(\alpha,H\alpha)}{(\alpha,\alpha)}$
从上面几节的内容你已经看出了，上述对欧空间中二次型的讨论对任意的酉空间的内积形式也是成立的。我们已经通过代数学给出了一个很直观的证明，后面变分法的内容会用其他方法证明。

本征值的极值性的另一个证明

令

$$J[\alpha] = \frac{(\alpha, H\alpha)}{(\alpha, \alpha)}$$

设 $J[\alpha]$ 在 $\alpha = \tilde{\alpha}$ 处取得极值，那么：
令

$$\alpha = \tilde{\alpha} + \varepsilon\beta, \varepsilon \in \mathbb{R}, \beta \in V$$

对于任意给定的 β，$J[\alpha]$ 都是 ε 的一个函数，并且应该在 $\varepsilon = 0$ 处取得极值。那么 $\frac{dJ[\varepsilon]}{d\varepsilon}\Big|_{\varepsilon=0} = 0$，现在可以做一下简单的运算并得到：

$$dJ|_{\varepsilon=0} = J[\tilde{\alpha} + \beta d\varepsilon] - J[\tilde{\alpha}] = \frac{1}{(\tilde{\alpha}, \tilde{\alpha})}\left[\left(\beta, (H - \tilde{\lambda})\tilde{\alpha}\right) + ((H - \tilde{\lambda})\tilde{\alpha}, \beta)\right] d\varepsilon$$

其中:

$$\tilde{\lambda} = \frac{(\tilde{\alpha}, H\tilde{\alpha})}{(\tilde{\alpha}, \tilde{\alpha})}$$

本征值的极值性的另一个证明（初始变分）

由于对于任意的 β，都有 $\left.\frac{dJ[\varepsilon]}{d\varepsilon}\right|_{\varepsilon=0}=0$，我们可以取

$$\beta=(H-\tilde{\lambda})\tilde{\alpha}$$

得到

$$H\tilde{\alpha}=\tilde{\lambda}\tilde{\alpha}$$

即：$J[\alpha]$ 在 H 的本征向量处取得极值，极小值是最小的本征值。

- 对于未归一化的向量，本质上就是增加了约束条件 (α,α) 等于 0，可以自己试着用拉格朗日不定乘子法证明。
- 对于其余的本征值，可以用不定乘子法证明。
- 上述证明过程不依赖 H 是有限维还是无穷维，也不依赖内积的形式，只要有内积的定义，H 是厄米算符。
- 最常用的地方是量子力学中，对于定义了内积的希尔伯特空间，所谓的力学平均值的“变分原理”就是运用了上述结论。

本征值的极小极大值性

量子力学中还有一类非常重要的泛函需要研究，为了体现普遍性，我下面采用 bra-ket 的内积记号：

Theorem (极小极大问题)

在约束 $\langle \alpha | \alpha \rangle = 1$ 下，对于满足 $\langle \beta | \alpha \rangle = 0$ 的 $|\alpha\rangle$，定义泛函

$$J[\beta] = \langle \alpha | H\alpha \rangle$$

这个泛函的极大值是 H 的第二个本征值 λ_1.

这个定理显然可以推广到一般，并且是十分容易得到证明的，只需证明 β 不取本征向量时，空间任意一个满足约束的向量作为 α 都使得泛函不大于 λ_n

利用初始变分方法导出欧拉方程

在各个学科中，我们经常遇到这样的一类泛函的极值问题：

$$J[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\,dx$$

其中 x_0,x_1,$y(x_0)$,$y(x_1)$ 为给定的，函数 F 足够的好。
这样的问题是最原始的，我们从他可以导出著名的欧拉方程。
这个问题是本征值的极值问题的一个显然推广，我们利用几乎完全类似的方法：
选取足够好的函数 $\eta(x)$, 满足 $\eta(x_0) = 0$,$\eta(x_1) = 0$ 并且设泛函 $J[y(x)]$ 在 $\tilde{y}$ 时取得极值，设

$$y(x) = \tilde{y}(x) + \varepsilon\eta(x)$$

此时泛函化作 ε 的函数，同样在 $\varepsilon = 0$ 处导数为 0 结果：

$$\int_{x_0}^{x_1} \eta(x)(F_y - \frac{d}{dx}F_{y'})\,dx = 0$$

变分几号的引入

上述讨论方法对多个自变量、高阶导数、一侧或双侧不固定的条件都是适用的，下去可以严格的证明他们。我们更加具体化刚刚的过程，可以引入更加方便的记号：

- 所谓的“$y(x)=\tilde{y}(x)+\varepsilon\eta(x)$”可以设 $\delta y=\varepsilon\eta(x)$ 为所谓的 y 的变分。
- 可以引入 $\delta J[\cdot]=\varepsilon J'(0)$(后一个 J 是函数而非泛函)

这些记号有着一些性质，你总可以从上述定义式子得出，下面列出额外的。

- 对已经确定的函数变分是 0，即：$\delta cste(x)=0$，这给出了最重要的一个变分：$\delta x=0$
- 变分符号和导数符号、积分符号可以交换位置。

e.g.

我们计算几个简单的变分来练练手:

Theorem (哈密顿方程的导出)

求泛函

$$S = \int_{t_1}^{t_2} (p\dot{q} - H(p,q,t))dt$$

等价的方程，边界是固定的

Theorem (稳定方程的导出)

求泛函

$$\iint_S \frac{1}{2}(u_x^2 + u_y^2)dxdy$$

等价的方程，边界是固定的

e.g.

Theorem (稳定方程的导出)

求泛函

$$J[\varphi] = \iint_R (\varphi_x^2 + \varphi_y^2 - 2\varphi g) dxdy$$

等价的方程，边界是固定的, 其中 g 是已知的函数。

Theorem (波动方程的导出)

求泛函

$$J = \int_{t_1}^{t_2} \int_0^l (\rho u_t^2 - \mu u_x^2) dxdt$$

等价的方程，边界是固定的.

泛函约束

Theorem (广义的等周问题)

在固定边值条件下，求一个 y，使得满足约束

$$K = \int_{x_1}^{x_2} G(x, y', y'')dx = c$$

并使得积分

$$J = \int_{x_1}^{x_2} F(x, y', y'')dx$$

取得极值

上述问题可以直接从原始变分方法获得解答，或者更加简洁地，利用乘子法的思想，构造一个新的泛函：

$$W = \int_{x_1}^{x_2} (F + \lambda G)dx$$

并且使之成为极型。

有限约束

这一类约束是在理论力学中最常看到的，我们将用理论力学常用的记号写下这个问题：

Theorem (有限约束问题)

在固定边值条件下，求一个 $q_i(t)$，使得满足约束

$$G(q_i) = 0$$

并使得积分

$$J = \int_{t_1}^{t_2} L(q'_i, q_i, t) dt$$

取得最小。

从分析学的角度看，这就是给定曲面上的曲线给出 J 极值的情况，一个简单的想法是从 L 消去一个自变量，下面直接给出结果：

$$\frac{d}{dt} L_{\dot{q}_i} - L_{q_i} + \lambda(t) \frac{\partial G}{\partial q_i} = 0$$

微分方程约束

事实上，上面讨论的约束问题都没有包括下面的非完整约束。

Theorem (有限约束问题)

在固定边值条件下，求一个 y，使得满足约束

$$G(x, y, y') = 0$$

并使得积分

$$J = \int_{x_1}^{x_2} F(x, y', y'')dx$$

取得极值，且 $G(x, y, y')$ 无法转化成任何不含导数的函数的微分。

在这里，我只提供结论：我们不得不引入拉格朗日乘函来解决这个问题，问题的详细证明是复杂的。

直接方法

至此，我们讨论了原始变分方法可以适用于大部分的变分问题，我们引入了形式变分记号，并就此讨论了几种约束（完整、非完整）情况下的变分原理。这里我们主要运用的思想是：尽量使用已有的分析学知识将之转化为函数的极值问题，并利用微积分中的“多元函数极值条件”来解决这类变分问题。

下面我们的目光将放在另一个思想上

Theorem (变分问题的直接解法思想)

变分问题的直接解法，总是在于造出极小化的序列来以这个序列为基础，通过极限过程来获得问题的解。

里茨方法

下面是里茨方法的思想步骤：

- 做出确定的、完备的（如果是正交的那就更可爱了），一组基函数 w_n
- 这组基函数通过线性组合成 φ_n,$\varphi_n = \sum_i c_i w_i$
- 这样的适当的线性组合，使得 $J[\varphi] - J[\varphi_n]$ 为任意小。
- 泛函极值问题成为 c_n 这 n 个系数的极值问题。

我们断言当 n 逼近无穷时问题总是有解的，这也是维尔斯特拉斯定理保证了的。
但是问题往往没有这么简答，对于单独的问题，这个方法十分依赖基函数的适当性，后面会给出例子

有限差分法

有限差分法其实是里茨方法的一个特例，

Theorem (有限差分法)

考虑形如

$$D[y] = \int_{x_1}^{x_2} F(x, y, y')dx$$

的单积分极小值问题，考虑每一个子区间都是线性的基函数，做一个 Ritz 方法，此变分问题化作：

$$\sum_{i=0}^{m} F(x_i, y_i, \frac{y_{i+1} - y_i}{\Delta x})\Delta x = \min$$

可以证明，上述差分方法在 m 取极限时，误差的缩小仅仅和 m 的倒数成比例。

直接方法的应用举例

Theorem (势函数)

区域

$$\Omega = \begin{cases} 0 \leqslant x \leqslant a \\ 0 \leqslant y \leqslant b \end{cases}$$

求泛函

$$D[\varphi(x,y)] = \iint_\Omega (\varphi_x^2 + \varphi_y^2) dxdy$$

的极值和极值函数。
φ 足够好，在边界上取值为 0，并且已经归一化（满足约束）：

$$H[\varphi] = \iint_\Omega \varphi^2 dxdy = 1$$

直接方法的应用举例

为了求上述泛函的极值，我们将函数用三角函数为基进行（傅里叶）展开。

$$\varphi = \sum_{m,n=1}^{\infty} c_{mn} \sin(\frac{m\pi x}{a}) \sin(\frac{n\pi y}{b})$$

代入待求泛函和约束泛函中，注意运用帕塞瓦尔等式化简，得到:

$$D = \frac{ab\pi^2}{4} \sum_{m,n=1}^{\infty} c_{mn}^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2} \right)$$

$$H = \frac{ab}{4} \sum_{m,n=1}^{\infty} c_{mn}^2 = 1$$

H 现在是 $c_m n$ 的函数，极值是显然的，只需要约束要求的 $c_{11} = \frac{\sqrt{ab}}{2}$，其余的系数都是 0。

直接方法的应用举例

因此变分问题的解是:

$$u = \frac{2}{\sqrt{ab}} \sin \frac{\pi x}{a} \sin \frac{\pi y}{b}$$

$$D[\varphi] \geqslant d = \pi(\frac{1}{a^2} + \frac{1}{b^2})$$

Theorem (圆区域的狄利克雷问题)

区域

$$\Omega : x^2 + y^2 \leqslant 1$$

求积分

$$D[\varphi] = \iint_{\Omega} (\varphi_x^2 + \varphi_y^2) dxdy$$

的极小

在 Ω 上应用极坐标 r, θ，从而将积分化作：

$$D[\varphi] = \int_0^{2\pi} d\theta \int_0^1 (\varphi_r^2 + \frac{\varphi_\theta^2}{r^2}) rdr$$

假定边值是一个傅里叶级数

$$\varphi(1,\theta)=f(\theta)=\frac{1}{2}a_0+\sum_{n=1}^{\infty}(a_n\cos n\theta+b_n\sin n\theta)$$

将 φ 做傅里叶展开:

$$\varphi=\frac{1}{2}f_0+\sum_{n=1}^{\infty}[f_n(r)\cos n\theta+g_n(r)\sin n\theta]$$

其中

$$f_n(1)=a_n, g_n(1)=b_n$$

按照三角函数组的完备性关系，我们可以将泛函积分出来，得到:

$$D[\varphi]=\pi\int_0^1[f_0'(r)]^2rdr+\pi\sum_{n=1}^{\infty}\int_0^1\{[f_n'(r)]^2+\frac{n^2}{r^2}[f_n(r)]^2\}rdr$$
$$+\pi\sum_{n=1}^{\infty}\int_0^1\{[g_n'(r)]^2+\frac{n^2}{r^2}[g_n(r)]^2\}rdr$$

直接方法的应用举例

由于上面的思想，我们只需要处理这些极小问题：

$$\int_0^1 \{[f_n{}'(r)]^2 + \frac{n^2}{r^2}[f_n(r)]^2\} r dr = \min$$

$$\int_0^1 \{[g_n{}'(r)]^2 + \frac{n^2}{r^2}[g_n(r)]^2\} r dr = \min$$

当 $r = 1$ 时，f_n 和 g_n 依次取得 a_n 和 b_n 将 f_n 和 g_n 在 $r = 0$ 处做泰勒展开，很简单的就可以得出答案（自己检验）原来变分问题的解为：

$$\varphi(r,\theta) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} r^n(a_n \cos n\theta + b_n \sin n\theta)$$

极小为：

$$D \geqslant d = \pi \sum_{n=1}^{\infty} n(a_n^2 + b_n^2)$$